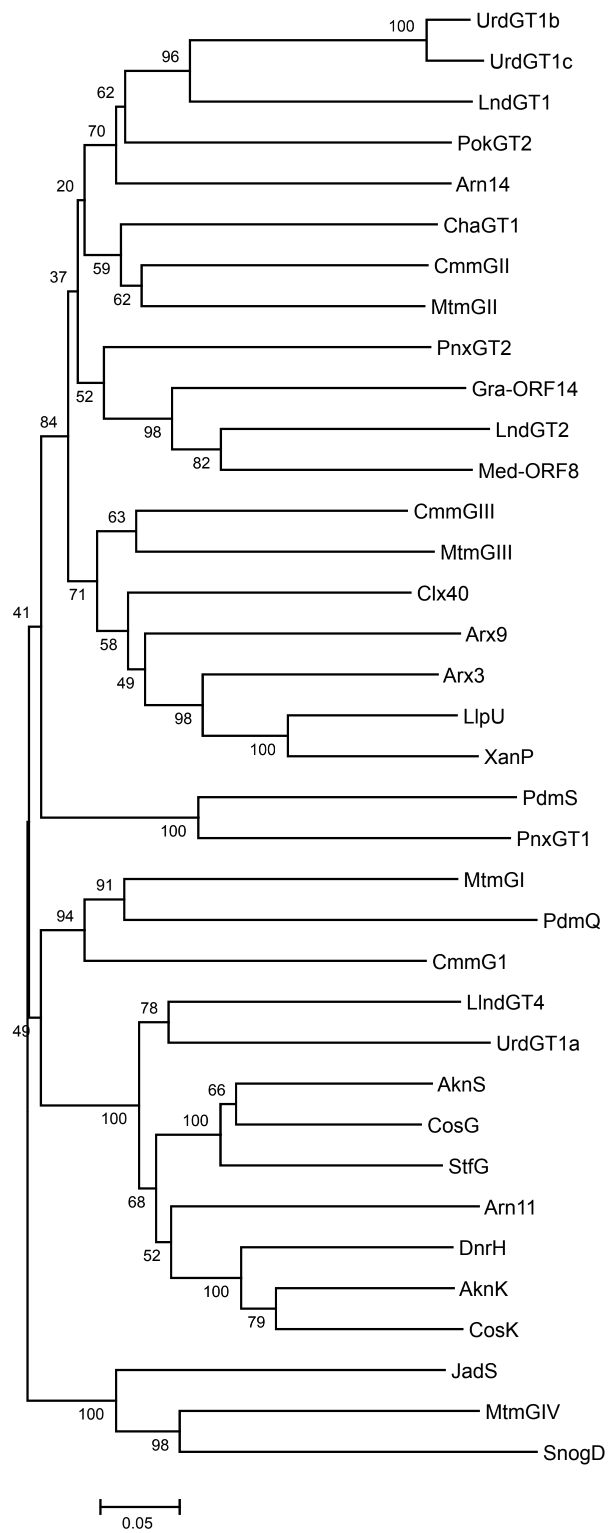
Supporting Information for

Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study

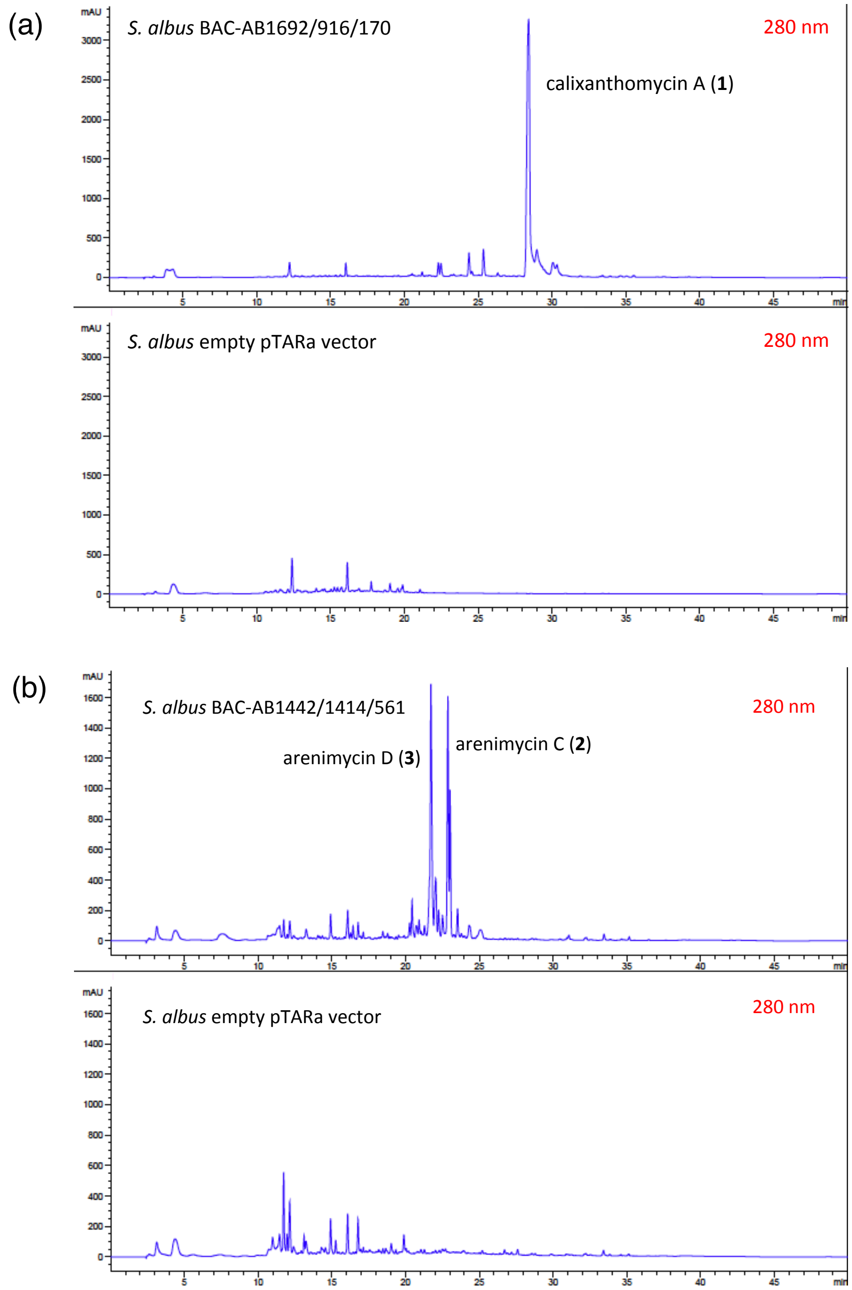
Hahk-Soo Kang and Sean F. Brady*

Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, New York, NY 10065, United States

Table of contents

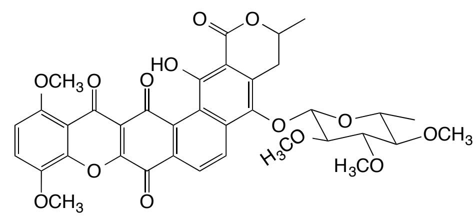


Supplementary figure 1. The maximum likelihood phylogenetic tree of glycosyltransferases genes from diverse type II PKS gene clusters.



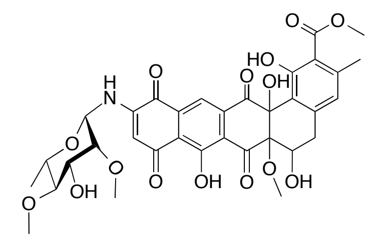
Supplementary figure 2. Comparison of the HPLC chromatograms: (a) HPLC chromatograms of the acid extracts of *S. albus* BAC-AB1692/916/170 and *S. albus* empty vector, (b) HPLC chromatograms of the neutral extracts of *S. albus* BAC-AB1442/1414/561 and *S. albus* empty vector.

(a)

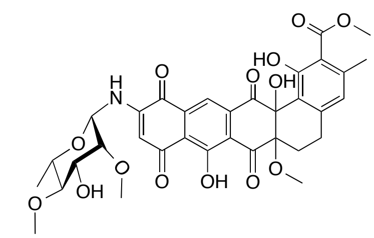


IB-00208

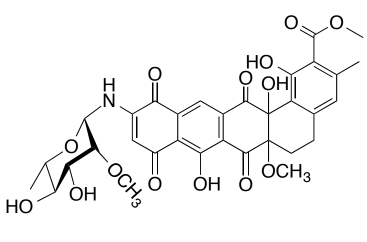
(b)



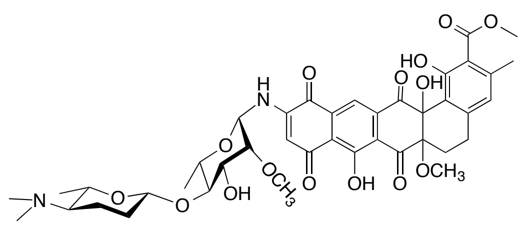
SF2446A1



SF2446B1



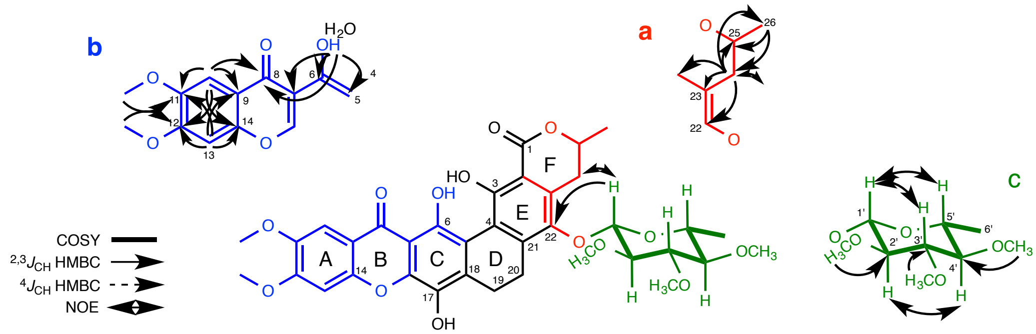
Arenimycin A



Arenimycin B

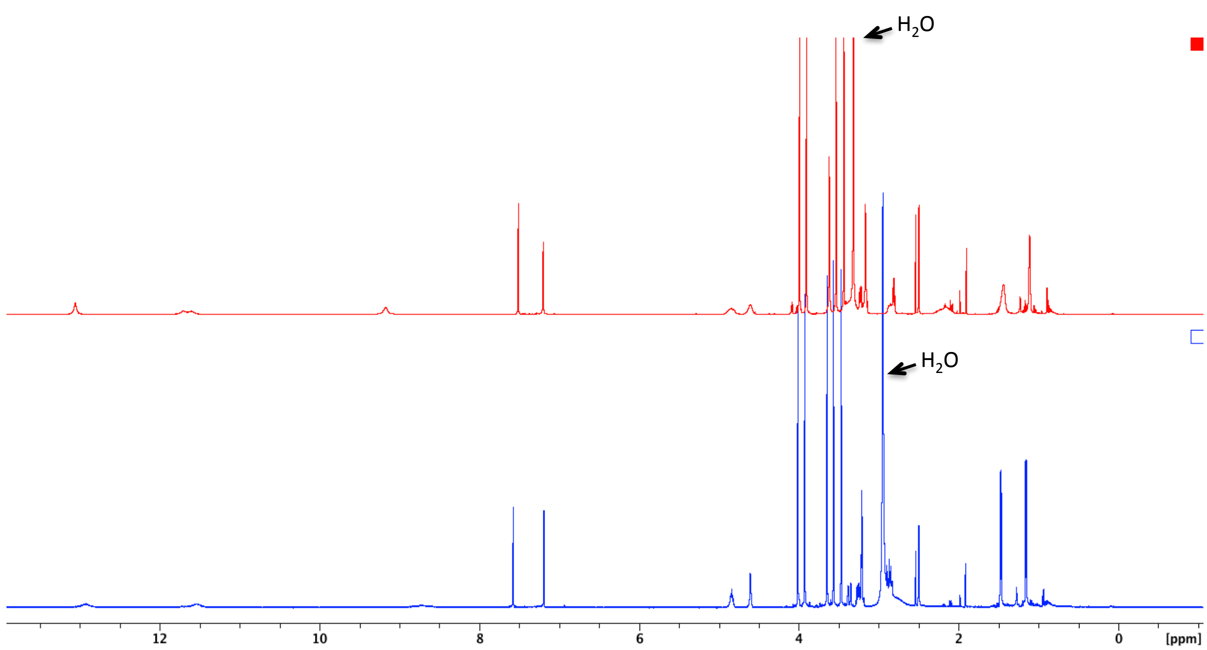
Supplementary figure 3. Sturctures of previously reported compounds that are related to (a) calixanthomycin A (**1**), and (b) arenimycins C (**2**) and D (**3**).[1]

Supplementary Discussion 1. Structure determination of calixanthomycin A (**1**).



Supplementary figure 4. Key 2D NMR correlations used to define the structure of calixanthomycin A (**1**).

The structure of **1** was determined by spectroscopic methods, including HRESIMS, and 1D and 2D NMR. A pseudo molecular ion at m/z 695.2327 ($[M+H]^+$) observed in the HRESIMS spectrum suggested a molecular formula of $C_{36}H_{39}O_{14}$. The ^1H NMR spectrum of **1** showed proton signals characteristic of a glycosylated aromatic polyketide, including signals for phenolic hydroxyl (10-15 ppm), aromatic (6-9 ppm), sugar anomeric (4-6 ppm), hydroxylated methine (2-5 ppm), methoxy (3-4 ppm) and methyl (1-2 ppm) protons (Supplementary table 1). Line broadening was observed for some signals (Supplementary figure 5). Line broadening was significantly reduced when the temperature was raised up to 100 °C, thus all the 2D and ^{13}C NMR spectra were recorded at this temperature.



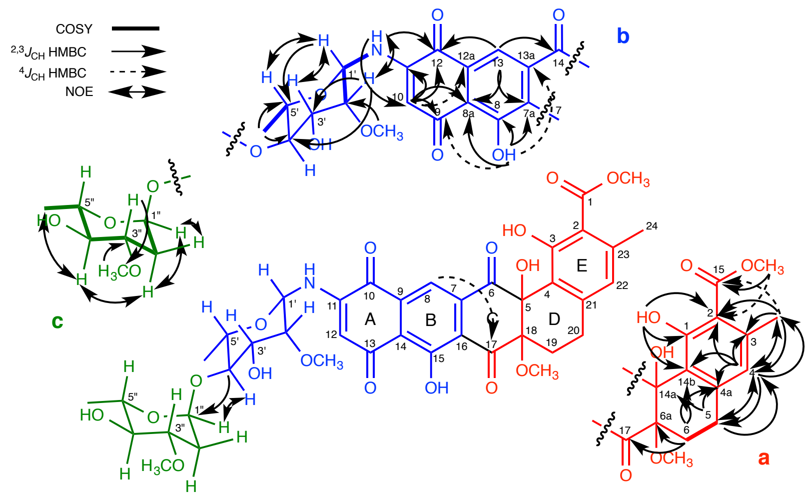
Supplementary figure 5. The ^1H NMR spectrum of **1** at room temp. (red) and at 100 ºC (blue).

Analysis of the 2D NMR spectrum of **1** including COSY, HSQC and HMBC established three sub-structures designated as **a**, **b** and **c** (Supplementary figure 4). HMBC correlations from H-26 (δ_H 1.49) to C-24 (δ_C 28.6) and C-25 (δ_C 75.0) and from H_2-

24 (δ_H 2.91 and 3.38) to C-2 (δ_C 106.2), C-22 (δ_C 141.0), C-23 (δ_C 131.1), C-25 (δ_C 75.0) and C-26 (δ_C 19.8) as well as COSY correlations between H-25 (δ_H 4.84) and H_3-26 (δ_H 1.49), and between H_2-24 (δ_H 2.91 and 3.38) and H-25 (δ_H 4.84) were used to define the substructure **a**. The second substructure (**b**) was also deduced from HMBC correlations. HMBC correlations from H-10 (δ_H 7.58) to C-8 (δ_C 179.9), C-9 (δ_C 112.3), C-12 (δ_C 156.1) and C-14 (δ_C 151.6), and from H-13 (δ_H 7.19) to C-9 (δ_C 112.3), C-11 (δ_C 146.7), C-12 (δ_C 156.1) and C-14 (δ_C 151.6) together with HMBC correlations from 11-OCH_3 (δ_H 3.93) and 12-OCH_3 (δ_H 4.01) to C-11 (δ_C 146.7) and C-12 (δ_C 156.1), respectively, indicated the presence of an ortho-dimethoxy substituted xanthone moiety. The last substructure (**c**) was defined as a tri-O-methylated sugar moiety based on COSY correlations between H-1' to H-6' as well as HMBC correlations from OCH_3 (δ_H 3.65) to C-2' (δ_C 83.6), OCH_3 (δ_H 3.57) to C-3' (δ_C 84.9) and OCH_3 (δ_H 3.47) to 4' (δ_C 58.8). NOE correlations observed between H-1' (δ_H 4.61), H-3' (δ_H 3.22) and H-5' (δ_H 3.27) and between H-2' (δ_H 3.22) and H-4' (δ_H 2.86) indicated the sugar moiety to be a tri-OMe-quinovose.

The substructures **a** and **c** were connected together based on an HMBC correlation from H-1' to C-22, which indicates that the tri-OMe-quinovose sugar moiety is attached to C-22 via an O-glycosidic linkage. No correlations were observed in the HMBC spectrum to connect the substructures **a** and **b**. Eight additional carbon signals (beyond those found in the substructures **a**, **b** and **c**) are seen in the ^{13}C NMR spectrum. This includes signals for one carbonyl (δ_C 169.1), five aromatic (δ_C 155.3, 142.3, 136.2, 131.9 and 119.4) and two aliphatic (δ_C 22.6 and 23.3) carbons. The two aliphatic carbons display faint negative 2D correlations with broad proton signals in the edited HSQC spectrum, indicating that these two signals represent methylene carbons. This is commonly observed for pentangular polyphenols with a single bond in the ring D due to the conformational interchange between two atropisomers.[2] In light of this observation, a typical pentangular polyphenol scaffold was assembled by connecting the D ring to the substructure **a** via C-3 (δ_C 155.3), C-4 (δ_C 119.4) and C-21 (δ_C 131.9), and to the substructure **b** via C-17 (δ_C 142.3) and C-18 (δ_C 136.2). Based on the molecular formula predicted by HRESIMS, the F ring was closed as a lactone through the carbonyl predicted by the last unassigned carbon signal we observed (C-1, δ_C 169.1). This structure satisfies all the 1D and 2D NMR, and HRESIMS data we collected and is easily rationalized biosynthetically based on the genes found in the clx gene cluster.

Supplementary discussion 2. Structure determination of arenimycins C (**2**) and D (**3**).



Supplementary figure 6. Key 2D NMR correlations used to define the structure of arenimycin C (**2**).

The structures of **2** and **3** were determined by spectroscopic methods, including HRESIMS, and 1D and 2D NMR. The HRESIMS spectrum of **2** displayed a pseudo molecular ion at *m/z* 812.2770 ($[M+H]^+$), suggesting the molecular formula of $C_{40}H_{45}NO_{17}$. The ^1H NMR spectrum of **2** displayed a signal distribution typical of a glycosylated aromatic polyketide, including signals for phenolic hydroxyl (10-15 ppm), aromatic (6-9 ppm), sugar anomeric (5-6 ppm), hydroxylated methine (2-5 ppm), methoxy (3-4 ppm) and methyl (1-2 ppm) protons (Supplementary table 2). Signal doubling was observed for some protons suggesting the presence of two isomers (Supplementary figure 7).

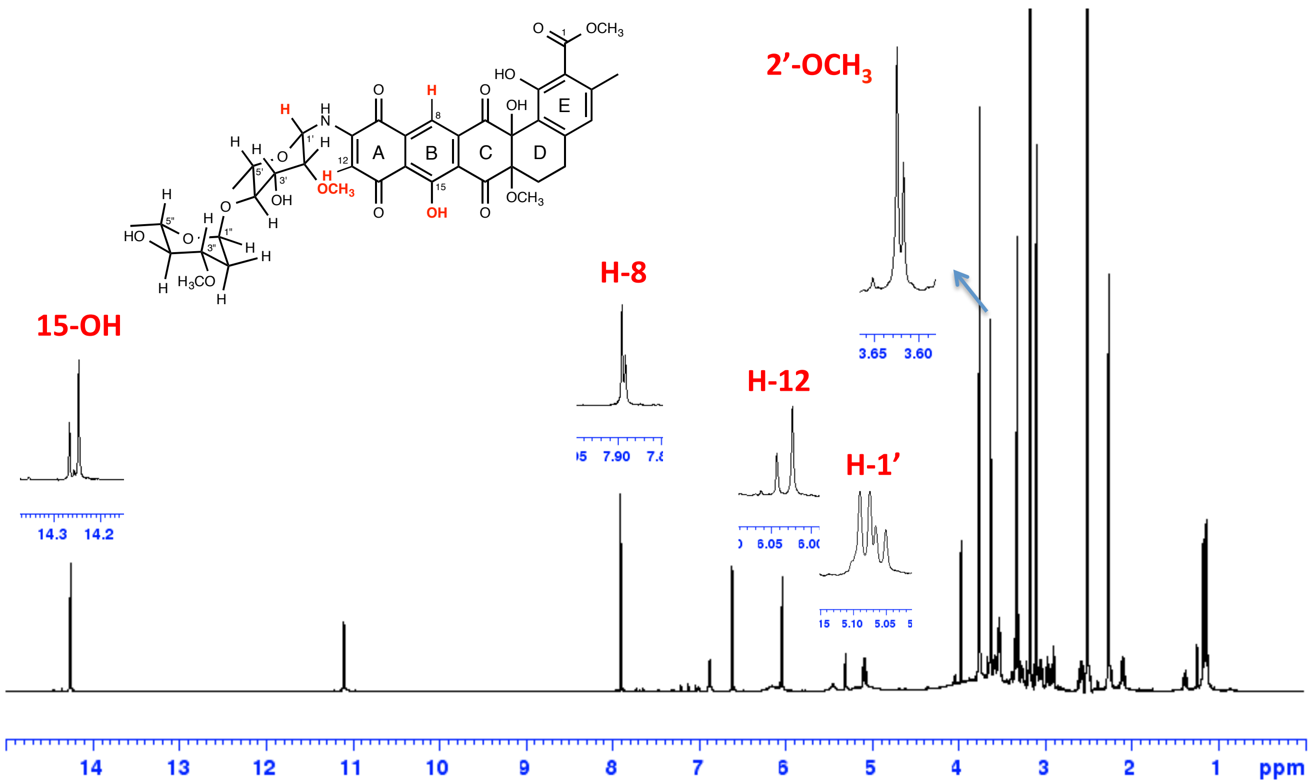
Analysis of the COSY, HMQC and HMBC spectra from **2** identified three substructures designated as **a**, **b** and **c** (Supplementary table 2 and Supplementary figure 6). The structure of a penta-substituted tetrahydronaphthalene (substructure **a**) was deduced using COSY correlations between H_2-20 and H_2-19 and HMBC correlations from H_2-20 to C-4, C-21 and C-22, from H-19 to C-5, C-17, C-18 and C-21 (ring D), from H-22 to C-2, C-4, C-20, C-21 and C-23, from H_3-24 to C-2, C-22 and C-23 and from 3-OH to C-2, C-3 and C-4 (ring E). An HMBC correlation from OCH_3 (Ac) to the C-1 carbonyl and a weak four-bond HMBC correlation from H_3-24 to C-1 placed an acetyl group at C-2.

The structure of a monoglycosylated tetra-substituted naphthaquinone (substructure **b**) was also determined using COSY and HMBC correlations. HMBC correlations from H-8 to C-16, C-14, C-10 and C-6, from 15-OH to C-14, C-15 and C-16, from H-12 to C-10, C-11, C-13 and C-14, and from NH to C-10 and C-12 together with weak four-bond correlations from 15-OH to C-7 and C-13, and from H-12 to C-9 established a tetra-substituted naphthaquinone substructure. This structure was further expanded to include a sugar moiety attached via an *N*-glycosidic linkage using COSY correlations (NH/H-1'/H-2' and H_3-6'/H-5') as well as HMBC correlations from H_3-6' to C-4' and from H-2' to C-3' and C-4'. An HMBC correlation from 2'-OCH_3 (δ_H 3.62) to C-2' placed a methoxy at C-2' and NOE correlations observed between H-1' and H-5', and H-3' and H-5' suggested the sugar was an *O*Me-rhamnose moiety.

The third substructure (substructure C) was determined to be an *O*-methylated deoxysugar based on sequential COSY correlations (H-1"/H_2-2"/H-3"/H-4"/H-5"/H_3-6") and an HMBC correlation from OCH_3 to C-3". An NOE correlation observed between H-2"$_{ax}$ and H-4" established this sugar as *O*Me-olivose. The substructures **a**, **b** and **c** were assembled through the use of HMBC and NOE correlation data (supplementary figure 6). A weak four-bond HMBC correlation from

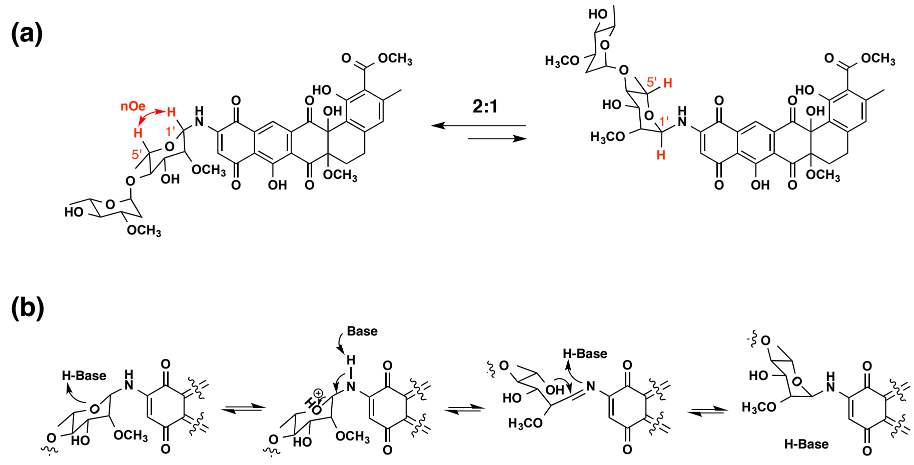
H-8 to C-17 was the only 2D correlation that bridged the substructures **a** and **b**. However, the creation of the para-quinone ring (C ring) is supported by chemical shifts observed for C-5 (δ_C 78.0), C-6 (δ_C 196.2), C-16 (δ_C 123.0) and C-17 (δ_C 190.1). Lastly, the substructure **c** (*O*Me-L-olivose) was attached to the substructure **b** via an *O*-glycosidic linkage between C-4' and C-1" based on an HMBC correlation from H-4' to C-1", to give the final structure of **2**.

The molecular formula of **3** was predicted to be $C_{32}H_{31}NO_{14}$ based on HRESIMS data (*m/z* = 654.1826, $[M+H]^+$). This differs from **1** by $C_8H_{14}O_3$ suggesting that **3** likely differed from **2** by the absence of one deoxy-sugar moiety. As predicted by the molecular formula, while NMR signals for the benzo[a]naphthacene quinone core structure are nearly identical between **2** and **3** (Supplementary table 2) significant differences were observed for NMR signals associated with the sugar moieties. Signals for the second sugar moiety and an *O*Me signal for the first sugar are absent in the NMR spectra of **3**, confirming that the structure of **3** is a monosaccharide analog of **2** bearing rhamnose instead of *O*Me-rhamnose.

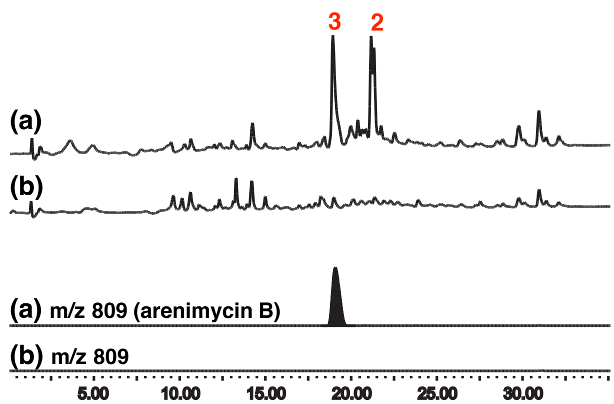


Supplementary figure 7. Signal doubling observed in the ^{1}H NMR spectrum of **2**.

In the ^{1}H NMR spectrum of **2**, a signal doubling was observed for some protons (ratio of 2:1), indicating the possibility that two interchangeable isomers are present in equilibrium. The same phenomenon was reported for the closely related SF2446 compounds.[1c,3] All of the doubled signals, which includes H-8, H-12, 15-OH, H-1' and 2'-OCH_3, were found in an interface region between the aglycone and the sugar moiety (supplementary figure 7). The signal doubling is likely caused by anomerization of the *N*-glycosidic linkage (supplementary figure 8). As would be expected from such an anomerization, in the NOE spectrum the major isomer shows an NOE correlation between H-1' and H-5' that is not seen in the minor isomer (Supplementary figure 8a). Attempts to individually characterize these two isomers have been so far unsuccessful.



Supplementary figure 8. Proposed anomerization of arenimycin C (**2**): (a) Compound **2** was isolated as what we predict is a mixture of two anomers in 2:1 ratio, (b) Our proposed mechanism of anomerization.



Supplementary figure 9. HPLC chromatograms (upper) and selective positive ion chromatograms (bottom) for m/z 809 (arenimycin B): (a) *S. albus* BAC-AB1442/1414/561 and (b) *S. albus* empty vector.

Supplementary protocol 1. The list of primers

- Degenerate primers used for library screening

KS_{α} degenerate primers

FW (5'-TSGCSTGCTTCGAYGCSATC-3')

RV (5'-TGGAANCCGCCGAABCCGCT-3')

KS_{β} degenerate primers

FW (5'-TTCGGSGGITTCCAGWSIGCSATG-3')

RV (5'-TCSAKSAGSGCSAISGASTCGTAICC-3')

- Primers used for TAR cloning

AB1692UPS_FW (5'-CTATCGATCTCGAGGCTGGACACCTGTCTCTACA-3'),

AB1692UPS_RV (5'-TCTACCGGAACAGTTAACTGTCAGATCCACCTGACTGC-3'),

AB170DWS_FW (5'- **GTTAAC**TGTTCCGGTAGAAGAAGC-3'),

AB170DWS_RV (5'-CCCTGCAGGAGCTCGTATCATCACTCCGCATTGTC-3').

AB1442UPS_FW (5'-CTATCGATCTCGAGGGATCAATGCCGTGGATCT-3'),

AB1442UPS_RV (5'-CAGTAGATGTTAACAGTTACGGGGAGGATGCGTA-3'),

AB561DWS_FW (5'-**GTTAAC**ATCTACTGTCGGTCGGGTT-3'),

AB561DWS_RV (5'-CCTGCAGGAGCTCGGTGGCTGTGTTGATGCCAGA-3').

Primers AB1692UPS_FW, AB170DWS_RV, AB1442UPS_FW and AB561DWS_RV include 15 bp sequences (underlined) that overlap with BmtI/SphI linearized pTARa capture vector. Primer AB1692UPS_RV and AB1442UPS_RV were designed to contain 15 bp overlaps (underlined) with the AB170DWS_FW and AB561DWS_FW primers and an HpaI site (bold), which was added to facilitate the linearization of the pathway-specific pTARa capture vector.

Supplementary table 1: NMR spectroscopic data for calixanthomycin A (**1**) at 100°C in DMSO-d_6

No.	calixanthomycin A (**1**)				
	δ_C[b]	δ_H[a]	*J* in Hz	HMBC	NOE
1	169.1				
2	106.2				
3	155.3				
4	119.4				
5	112.4				
6	150.2				
7	106.0				
8	179.9				
9	112.3				
10	104.8	7.58	s	8, 9, 11, 12, 14	11'-OCH_3
11	146.7				
12	156.1				
13	99.9	7.19	s	8, 9, 11, 12, 14	12-OCH_3
14	151.6				
16	144.3				
17	142.3				
18	136.2				
19	22.6	2.86	overlapped		
20	23.3	2.86	overlapped		
21	131.9				
22	141.0				
23	131.1				
24	28.6	2.91	dd (17.3, 10.1)	2, 22, 23, 25, 26	26
		3.38	dd (17.3, 3.6)		25, 26, 1'
25	75.0	4.84	m		24
26	19.8	1.49	d (6.3)	24, 25	24, 25
11-OCH_3	55.8	3.93	s	11	10
12-OCH_3	56.1	4.01	s	12	13
3-OH		8.74	brs		
6-OH		12.93	brs		
17-OH		11.54	brs		
1'	103.9	4.61	d (7.1)	2', 3', 5', 22	24, 3', 2'-OCH3
2'	83.6	3.22	overlapped	1'	2'-OCH_3
3'	84.9	3.22	overlapped	1'	3'-OCH_3
4'	84.2	2.86	t (9.5)	3', 5', 6', 4'-OCH_3	4'-OCH_3
5'	69.8	3.27	dd (9.5, 6.2)	1'	
6'	16.9	1.17	d (6.2)	4' 5'	4', 4'-OCH_3
2'OCH_3	59.4	3.65	s	2'	1', 2'
3'OCH_3	58.9	3.57	s	3'	3'
4'OCH_3	58.8	3.47	s	4'	4' 6'

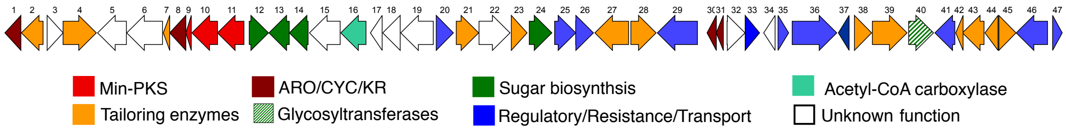
[a]recorded at 600 MHz, [b]recorded at 150 MHz, [a,b]signals were referenced to the DMSO-d_6 solvent signals (δ_H 2.50 and δ_H 39.51).

Supplementary table 2: NMR spectroscopic data for arenimycins C (**2**) and D (**3**) in DMSO-d_6

No.	arenimycin C (**2**)					No.	arenimycin D (**3**)		
	δ_C[b]	δ_H[a]	*J* in Hz	HMBC	NOE		δ_C[b]	δ_H[a]	J in Hz
1	170.1					1	170.0		
2	112.4					2	112.5		
3	156.4					3	156.3		
4	122.2					4	122.2		
5	78.0					5	78.0		
6	196.2					6	196.2		
7	140.9					7	140.8		
8	114.8	7.89	s	7a, 8a, 12, 14		8	114.7	7.89	s
9	133.9					9	133.9		
10	179.5					10	179.7		
11	147.8					11	148.3		
12	102.5	6.02	s	8, 8a, 9, 11, 12	1’, 5’	12	102.5	5.98	s
13	188.5					13	188.4		
14	118.1					14	118.3		
15	161.3					15	161.3		
16	123.0					16	122.9		
17	190.1					17	190.1		
18	87.7					18	87.7		
19	17.9	2.09 2.57	dt (12.5, 9.4) dd (12.5, 9.4)	4a, 5, 6a, 7, 14a	6a-OCH3	19	17.9	2.09 2.57	dt (12.6, 9.2) dd (12.6, 9.2)
20	26.1	2.94 3.05	dd (18.8, 9.4) dd (18.8, 9.4)	4, 4a, 6, 6a, 14b	4	20	26.1	2.95 3.05	dt (19.1, 9.2) dd (19.1, 9.2)
21	143.4					21	143.4		
22	123.2	6.61	s	1, 2, 3-CH_3, 5, 14b, 15	3-CH_3, 5	22	123.1	6.61	s
23	140.2					23	140.1		
24	22.0	2.26	s	1, 2, 3, 4, 15		24	22.0	2.26	s
18OCH$_3$	51.7	3.09	s	6a	6	18OCH$_3$	51.7	3.09	s
1-Ac	52.4	3.75	s	2, 15		1-Ac	52.4	3.74	s
NH		6.87	d (9.1)	10, 12	1’, 5’	NH		7.20	d (8.6)
3-OH		11.09	s	1, 2, 14b		3-OH		11.06	s
15-OH		14.25	s	7a, 8, 8a		15-OH		14.30	s
1’	78.1	5.08	d (9.1)	11, 2’, 5’	10, 3’, 5’	1’	78.9	4.94	d (8.6)
2’	80.7	3.52	overlapped			2’	69.9	3.74	overlapped
3’	74.5	3.75	overlapped			3’	73.2[c]	3.38	overlapped
4’	77.5	3.34	overlapped			4’	73.1[c]	3.17	overlapped
5’	71.5	3.52	overlapped	3’, 4’	10, 1’, 6’, 4”	5’	71.5	3.37	overlapped
6’	18.3	1.17	d (5.9)	4’, 5’	5’	6’	17.9	1.15	d (6.2)
2’OCH$_3$	61.5	3.62	s	2’					
1”	97.9	5.30	brd (3.5)	3”, 5”	2”, 3”-OCH$_3$				
2”	34.6	1.37 2.23	td (12.9, 3.8) dd (12.9, 4.9)	1”, 3”, 4”					
3”	77.7	3.28	overlapped	3”-OCH$_3$					
4”	75.5	2.89	t (9.0)	3”, 5”, 6”					
5”	68.4	3.57	m						
6”	17.7	1.12	d (6.2)	4”, 5”					
3”OCH$_3$	56.6	3.32	s	3”					

[a]recorded at 600 MHz, [b]recorded at 150 MHz, [a,b]signals were referenced to the DMSO-d_6 solvent signals (δ_H 2.50 and δ_H 39.51), [c]could be interchanged.

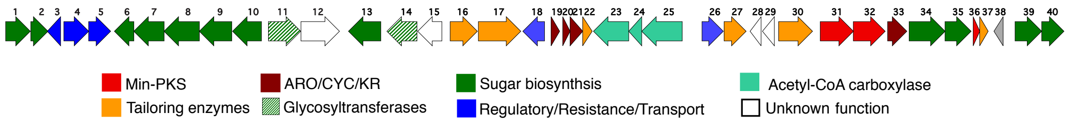
Supplementary table 3: Gene annotation table for the eDNA-derived Clx (calixanthomycin) gene cluster (GenBank KM881706)



Gene	Size (bp)	Proposed function	Homologous gene	Origin	Identity/ Similarity (%)	Accession NO.
ORF1	654	Hypothetical protein		*Actinoplanes friuliensis*	66/77	WP_023361877.1
ORF2	3,399	B12-dependent methionine synthase		*Actinoplanes friuliensis*	87/93	WP_023364409.1
ORF3	198	Methionine synthase		*Actinomadura madurae*	74/82	WP_021593142.1
ORF4	219	Hypothetical protein		*Kitasatospora* sp. NRRL B-11411	42/54	WP_030461720.1
ORF5	993	Ribokinase		*Saccharomonospora marina*	59/72	WP_009155260.1
Clx1	759	ketoreductase	pnxG	*Streptomyces* sp. TA-0256	65/80	BAJ52686
Clx2	1,029	*O*-methyltransferase	sanN	*Streptomyces* sp. SANK 61196	58/69	ADG86324.1
Clx3	756	monooxygenase	sanL	*Streptomyces* sp. SANK 61196	39/60	ADG86321.1
Clx4	1,578	FAD-binding monooxygenase	pnxO6	*Streptomyces sp.* TA-0256	43/57	BAJ52699
Clx5	1,365	Carbamoyl-phosphate synthase large subunit		*Catenulispora acidphila*	69/80	WP_012785313.1
Clx6	1,695	Acetyl-CoA carboxyl transferase		*Streptomyces catenulae*	65/77	WP_030283707.1
Clx7	306	monooxygenase	sanQ	*Streptomyces* sp. SANK 61196	50/60	ADG86327.1
Clx8	768	Bifunctional cyclase/monooxygenase	BenH	*Streptomyces sp.* A2991200	52/64	CAM58801.1
Clx9	258	Acyl carrier prtein (ACP)	Arx18	*Uncultured bacterium*	41/59	AHX24703.1
Clx10	1,215	Beta-ketoacyl synthase beta	XanE	*Streptomyces flavogriseus*	65/74	ADE22314.1
Clx11	1,257	Beta-ketoacyl synthase alpha	XanF	*Streptomyces flavogriseus*	79/87	ADE22315
Clx12	912	UDP-glucose 4-epimerase		*Amycolatopsis mediterranei*	56/65	WP_014467264.1
Clx13	960	dTDP-glucose 4,6-dehydratase		*Kutzneria* sp. 744	71/80	EWM15838.1
Clx14	891	Glucose-1-phosphate thymidylyltransferase		*Actinopolyspora halophila*	74/85	WP_017976869.1
Clx15	1,455	S-adenosyl-L-homocysteine hydrolase		*Thermobifida fusca*	84/90	WP_011292928.1
Clx16	1,197	S-adenosylmethionine synthetase		*Streptomyces scabrisporus*	83/89	WP_026218485.1
Clx17	555	Hypothetical protein		*Gordonia paraffinivorans*	45/65	WP_006902402.1
Clx18	828	Putative phosphoenol pyruvate synthase		Chondromyces crocatus	29/43	CAJ46695.1
Clx19	1,542	Spore coat protein A		*Amycolatopsis orientalis*	62/70	WP_016333765.1
Clx20	813	LuxR family transcriptional regulator		*Actinomadura madurae*	61/71	WP_021597215.1
Clx21	1,062	O-methyltransferase	Arx12	*Uncultured bacterium*	45/58	AHX24697.1
Clx22	1,470	Hypothetical protein		*Actinomadura rifamycini*	55/68	WP_026401562.1
Clx23	747	Methyltransferase type_12		*Amycolatopsis mediterranei*	58/73	KDO04566.1
Clx24	1,071	NAD-dependent dehydratase		*Streptomyces* sp. BoleA5	61/71	WP_020134963.1
Clx25	1,002	ABC transporter		*Geodermatophilaceae bacterium URHB0048*	64/74	WP_029337833.1
Clx26	843	ABC transporter	Arx7	*Uncultured bacterium*	52/70	AHX24692.1
Clx27	1,599	FAD-binding monooxygenase	pnxO4	*Streptomyces* sp. TA-0256	51/59	BAJ52694.1
Clx28	1,197	Putative cytochrome P450	pnxO5	*Streptomyces* sp. TA-0256	61/72	BAJ52696.1
Clx29	1,902	SARP family transcriptional regulator		*Streptomyces bingchenggensis*	46/59	WP_014174360.1
Clx30	441	Cupin (cyclase)		*Streptomyces* sp. CNT318	70/76	WP_027759557.1
Clx31	336	Cyclase	LlpCIII	*Streptomyces tendae*	76/86	CAM34347.1
Clx32	852	Hypothetical protein		*Streptomyces* sp. AW19M42	37/52	WP_024491883.1
Clx33	684	PadR family transcriptional regulator		*Thermomonospora curvata*	43/61	WP_012853995.1
Clx34	540	Hypothetical protein		*Nocardiopsis dassorvillei*	27/46	WP_013153792.1
Clx35	504	MarR family transcriptional regulator		*Chlorogloeopsis fritschii*	43/57	WP_026087606.1
Clx36	2,115	Lipoprotein		*Streptomyces* sp. NRRL F-5639	53/68	WP_031032030.1
Clx37	510	Acetyl-CoA carboxylase		*Streptomyces cattleya*	49/57	WP_014145734.1
Clx38	813	SAM-dependent methyltransfer-		*Amycolatopsis mediterranei*	53/64	WP_013227237.1

		ase				
Clx39	1,644	Putative FAD-binding monooxygenase	XanO5	*Streptomyces flavogriseus*	51/63	ADE22302.1
Clx40	1,179	Glycosyltransferase		*Actinoplanes* sp. N902-109	48/61	WP_015621193.1
Clx41	945	Hypothetical protein	Arx11	*Uncultured bacterium*	35/51	AHX24696.1
Clx42	348	Monooxygenase	SanQ	*Streptomyces* sp. SANK61196	38/54	ADG86327.1
Clx43	1,023	O-methyltransferase	SanN	*Streptomyces* sp. SANK61196	50/66	ADG86324.1
Clx44	618	Putative hydroxylase	pnxE2	*Streptomyces* sp. TA-0256	54/71	BAJ52685.1
Clx45	813	Putative dehydrogenase	pnxO1	*Streptomyces* sp. TA-0256	51/62	BAJ52671.1
Clx46	1,485	Peptide transporter		*Amycolatopsis mediterranei*	56/70	WP_013227259.1
Clx47	456	Hypothetical protein (monooxygenase)	pnxE1	*Streptomyces* sp. TA-0256	51/67	BAJ52672.1
ORF53	1,341	Bilirubin oxidase		*Micromonospora parva*	57/71	WP_030329397.1
ORF54	627	Hypothetical protein		*Galdieria sulphuraria*	43/60	XP_005707410.1
ORF55	675	Hypothetical protein		*Marmoricola aequoreus*	61/76	WP_030486102.1
ORF56	804	Photosystem reaction center subunit H		*Actinoplanes* sp. SE50/110	51/63	WP_014690426.1
ORF57	1,302	Putative nucleoside transporter yegT		*Arcticibacter svalbardensis*	68/85	WP_016194351.1

Supplementary table 4: Gene annotation table for the eDNA-derived Arn (arenimycin) gene cluster (GenBank KJ440489)



Gene	Size (bp)	Proposed function	Homologous gene	Origin	Identity/ Similarity (%)	Accession NO.
ORF1	627	Hypothetical protein		*Micromonospora sp.* ATCC 39149	68/78	WP_007074463.1
ORF2	510	Regulatory P domain-containing protein		*Amycolatopsis orientalis* HCCB10007	49/65	YP_008012359.1
ORF3	792	Regulatory P domain-containing protein		*Verrucosispora maris* AB-18-032	41/62	YP_004407220
ORF4	1,491	Hypothetical protein		*Longispora albida*	51/62	WP_018350322.1
ORF5	645	Threonine dehydratase		*Amycolatopsis sp.* ATCC 39116	53/67	WP_020420087.1
Arn1	945	dNDP-glucose synthase		*Micromonospora griseorubida*	75/84	BAC57039.1
Arn2	657	NDP-hexose 3,5-epimerase	RhaC	*Streptomyces olivaceus*	59/77	CAP11386.1
Arn3	510	MarR family transcriptional regulator		*Stackebrandtia nassauensis DSM 44728*	46/55	YP_710821.1
Arn4	951	ABC transporter ATP-binding protein		*Actinosynnema mirum* DSM 43827	63/74	YP_003098925.1
Arn5	843	ABC transporter		*Actinosynnema mirum DSM 43827*	51/71	YP_003098924
Arn6	744	NDP-hexose N,N-dimethyltransferase	SpnS	*Saccharopolyspora spinosa*	60/76	AAG23280.1
Arn7	1,161	Aminotransferase	SpnR	*Saccharopolyspora spinosa*	67/80	AAG23279.1
Arn8	1,353	NDP-hexose 3,4-dehydratase	RdmI	*Streptomyces purpurascens*	78/86	AAL24451.1
Arn9	1,242	Cytochrome P450	AntT	*Streptomyces galilaeus*	37/47	AAF73456.1
Arn10	1,107	Sugar O-methyltransferase	mra2	*Streptosporangium amethystogenes*	43/56	BAM98963.1
Arn11	1,236	Glycosyltransferase	EryclII	*Saccharopolyspora erythraea*	47/64	2YJN_A
Arn12	1,473	Hypothetical protein		*Actinomadura madurae*	51/67	WP_021596632.1
Arn13	1,236	Cytochrome P450		*Streptomyces avermitilis* MA-4680	41/55	NP_823237.1
Arn14	1,152	Glycosyltransferase	CloM	*Streptomyces roseochromogenes* subsp. Oscitans DS 12.976	48/59	AAN65229.1
Arn15	924	Aldo/keto reductase		*Streptomyces chartreusis*	51/66	WP_010041698.1
Arn16	1,041	O-methyltransferase	Lcz35	*Streptomyces sanglieri*	48/62	ABX71152.1
Arn17	1,623	FAD-dependent monooxygenase	GrhO8	*Streptomyces sp.* JP95	53/67	AAM33675.1
Arn18	831	SARP family transcriptional regulator		Frankia sp. QA3	41/54	WP_009739936.1
Arn19	327	Cyclase	sanD	*Streptomyces sp.* SANK61196	59/75	ADG86313.1
Arn20	282	Cupin (cyclase)		*Micromonospora sp.* ATCC 39149	60/69	WP_007071329.1
Arn21	486	Aromatase	WhiE VI	*Streptomyces gancidicus*	66/77	WP_006134520.1
Arn22	363	monooxygenase	XanO6	*Streptomyces flavogriseus*	57/67	ADE22309.1
Arn23	1,350	Carbamoyl-phophate synthase large subunit		*Streptomyces scabrisporus*	67/79	WP_020553758.1
Arn24	498	Biotin carboxyl carrier protein	TamJ	*Uncultured bacterium*	63/80	AFY23041.1
Arn25	1,551	Acetyl-CoA carboxyltransferase subunit beta		*Frankia sp.* BCU110501	59/71	WP_018501318.1
Arn26	816	SARP family transcriptional regulator	tsuY	*Streptomyces tsukubaensis*	52/66	CBY91988.1
Arn27	843	Methyltransferase	ChaI	*Streptomyces chartreusis*	43/56	CAH10176.1
Arn28	420	Hypothetical protein		*Actinoplanes friuliensis* DSM 7358	45/57	YP_008735210.1
Arn29	456	Hypothetical protein	RubQ	*Streptomyces collinus*	49/64	AAM97373.1
Arn30	1,296	Cytochrome P450		*Streptomyces sp.* R1-NS-10	43/56	WP_019071127.1
Arn31	1,269	Beta-ketoacyl synthase alpha	PdmA	*Actinomadura hibisca*	74/82	BAA23144.1
Arn32	1,218	Beta-ketoacyl synthase beta	PdmB	*Actinomadura hibisca*	65/74	BAA23145.1
Arn33	738	3-oxoacyl-ACP reductase	pnxG	*Streptomyces sp.* TA-0256	62/75	BAJ52686.1
Arn34	1,377	NDP-hexose 2,3-dehydratase	ssfS3	*Streptomyces sp.* SF2575	53/70	ADE34510.1
Arn35	999	NDP-hexose 3-ketoreductase	slgS4	*Streptomyces lydicus*	59/70	CBA11563.1
Arn36	255	Acyl carrier protein	ncnC	*Streptomyces arenae*	54/70	AAD20269.1
Arn37	324	Monooxygenase	PdmH	*Actinomadura hibisca*	58/72	ABM21754.1
Arn38	351	Transposase IS110		*Salinispora pacifica*	89/90	WP_018818993.1
Arn39	1,020	NDP-hexose 4,6-dehydratase		*Streptomyces sp.* SirexAA-E	74/85	YP_004802450.1
Arn40	852	NDP-hexose 4-ketoreductase		*Actinoplanes friuliensis* DSM 7358	81/86	YP_008734543.1
ORF46	348	DNA-directed DNA polymerase		*Frankia alni* ACN14a	55/65	YP_712925.1
ORF47	564	Recombinase		*Actinoplanes sp.* N902-109	52/68	YP_007950555.1
ORF48	405	Transposase, partial		*Streptomyces filamentosus*	67/75	WP_006125260.1

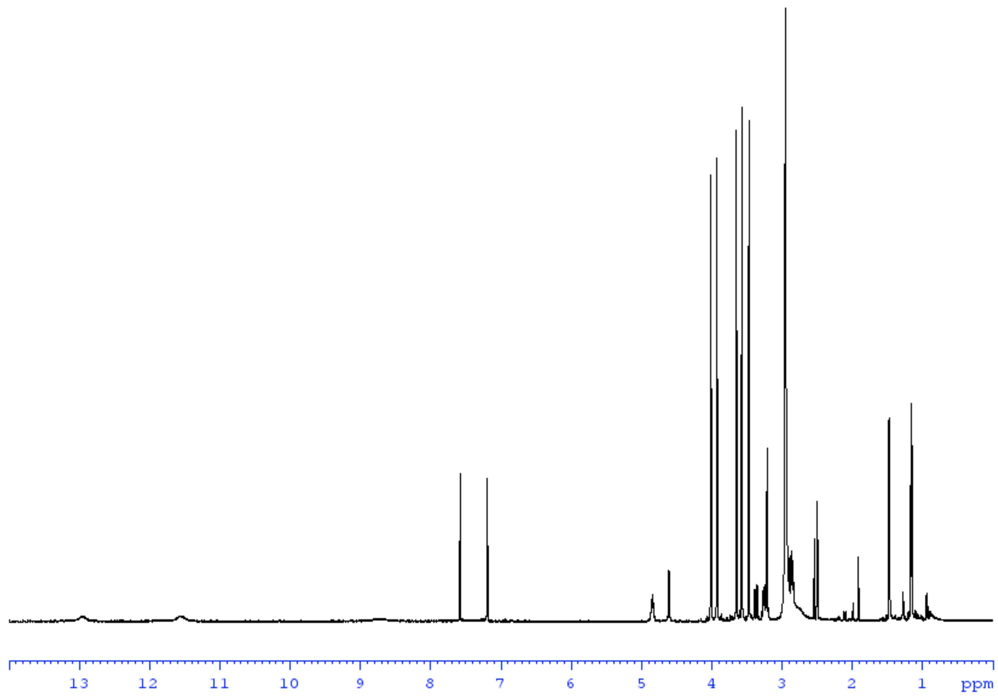
ORF49	417	Transposase		*Streptomyces violaceoruber*	62/79	NP_862176.1
ORF50	978	ATP dependent DNA ligase		*Actinoplanes sp.* SE50/110	43/57	YP_006268876.1

References for supplementary material:

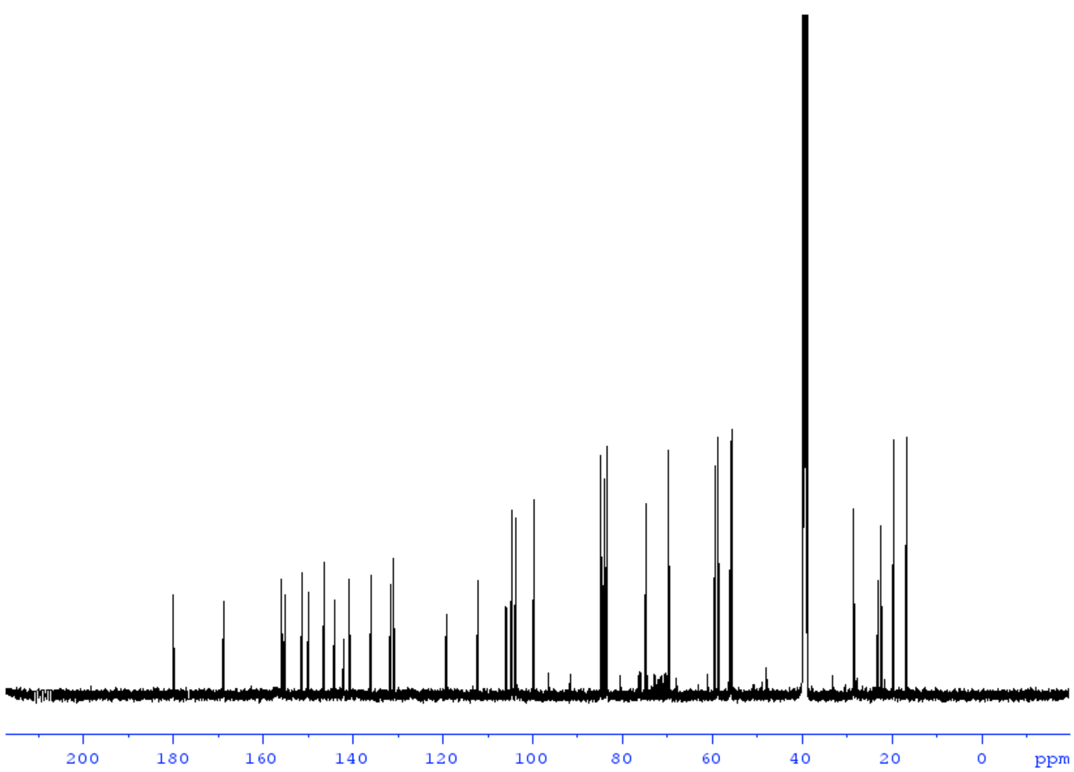
(1) (a)Asolkar, R. N.; Kirkland, T. N.; Jensen, P. R.; Fenical, W. *J. Antibiot.* **2010**, *63*, 37. (b) Gomi, S.; Sasaki, T.; Itoh, J.; Sezaki, M. *J. Antibiot.* **1988**, *41*, 425. (c) Kersten, R. D.; Ziemert, N.; Gonzalez, D. J.; Duggan, B. M.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. *Proc Natl Acad Sci U S A* **2013**, *110*, E4407.(d)Malet-Cascon, L.; Romero, F.; Espliego-Vazquez, F.; Gravalos, D.; Fernandez-Puentes, J. L. *J. Antibiot.* **2003**, *56*, 219.

(2) Eguchi, T.; Kondo, K.; Kakinuma, K.; Uekusa, H.; Ohashi, Y.; Mizoue, K.; Qiao, Y. F. *J. Org. Chem.* **1999**, *64*, 5371.

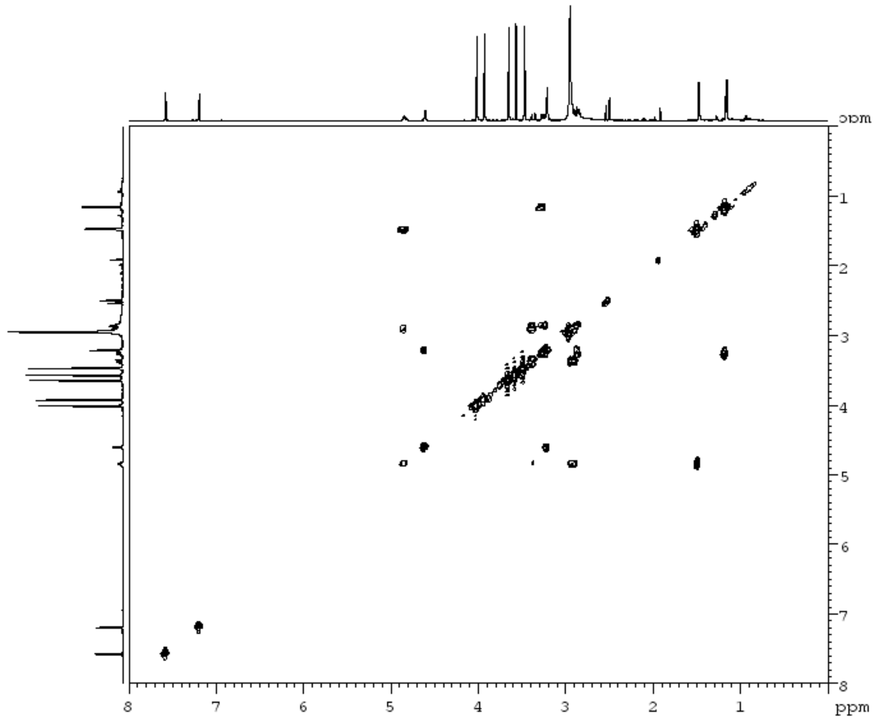
(3) (a)Asolkar, R. N.; Kirkland, T. N.; Jensen, P. R.; Fenical, W. *J. Antibiot.* **2010**, *63*, 37. (b) Gomi, S.; Sasaki, T.; Itoh, J.; Sezaki, M. *J. Antibiot.* **1988**, *41*, 425. (c) Takeda, U.; Okada, T.; Takagi, M.; Gomi, S.; Itoh, J.; Sezaki, M.; Ito, M.; Miyadoh, S.; Shomura, T. *J. Antibiot.* **1988**, *41*, 417.



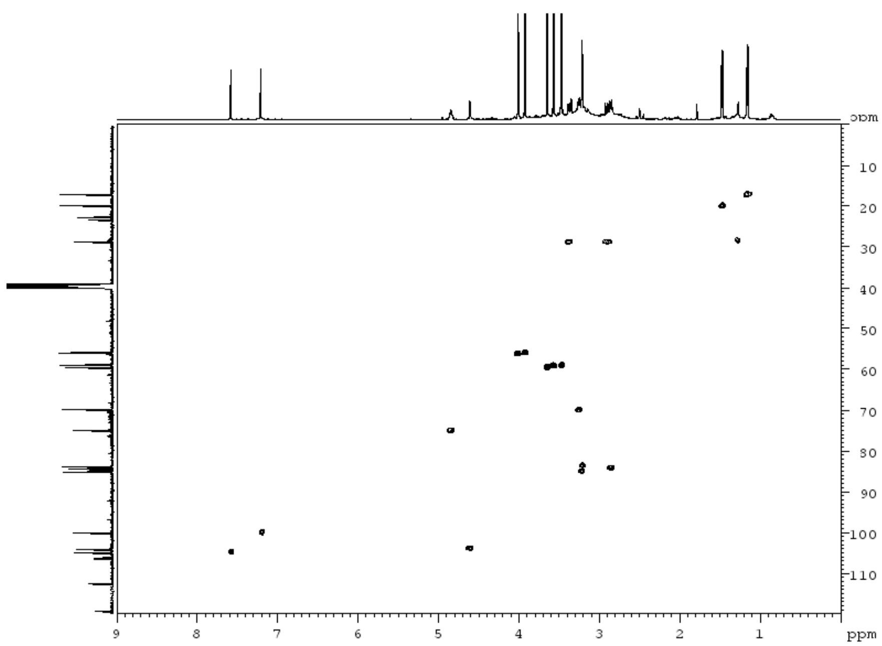
Supplementary figure 10. H NMR spectrum (DMSO-d_6, 500 MHz) of calixanthomycin A (**1**) at 100 °C.



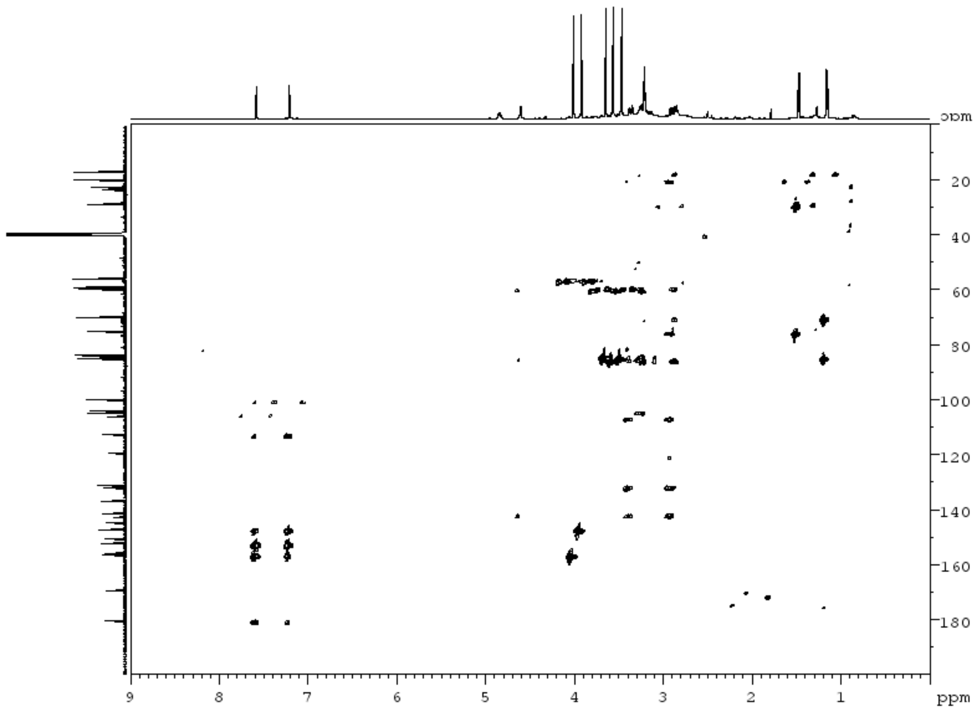
Supplementary figure 11. ^{13}C NMR spectrum (DMSO-d_6, 125 MHz) of calixanthomycin A (**1**) at 100 °C.



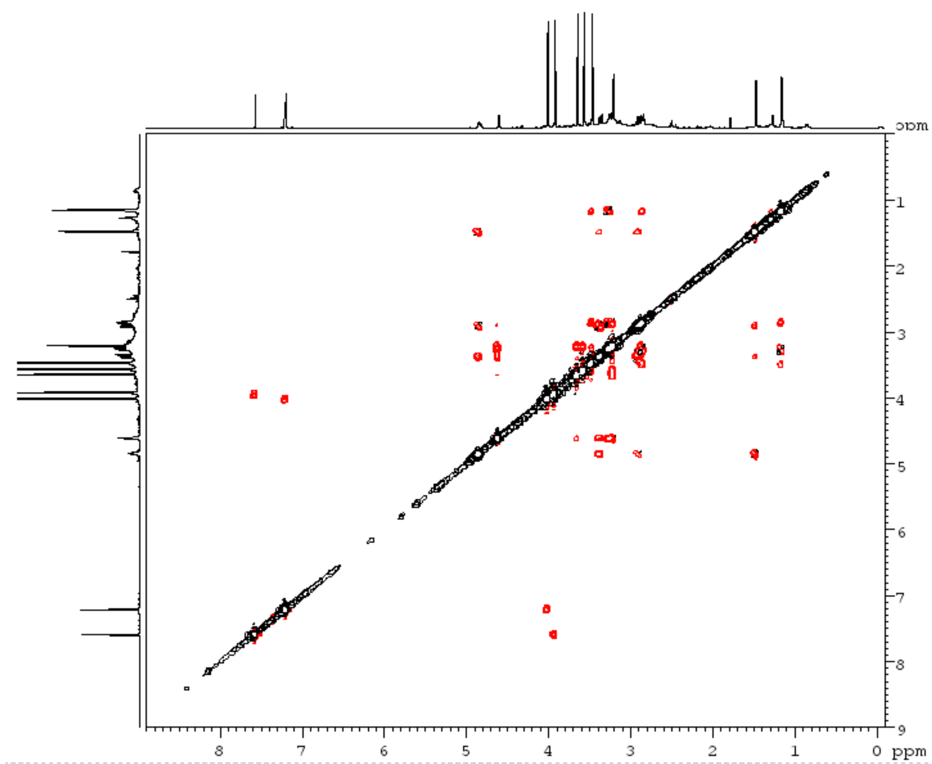
Supplementary figure 12. COSY spectrum (DMSO-d_6, 500 MHz) of calixanthomycin A (**1**) at 100 °C.



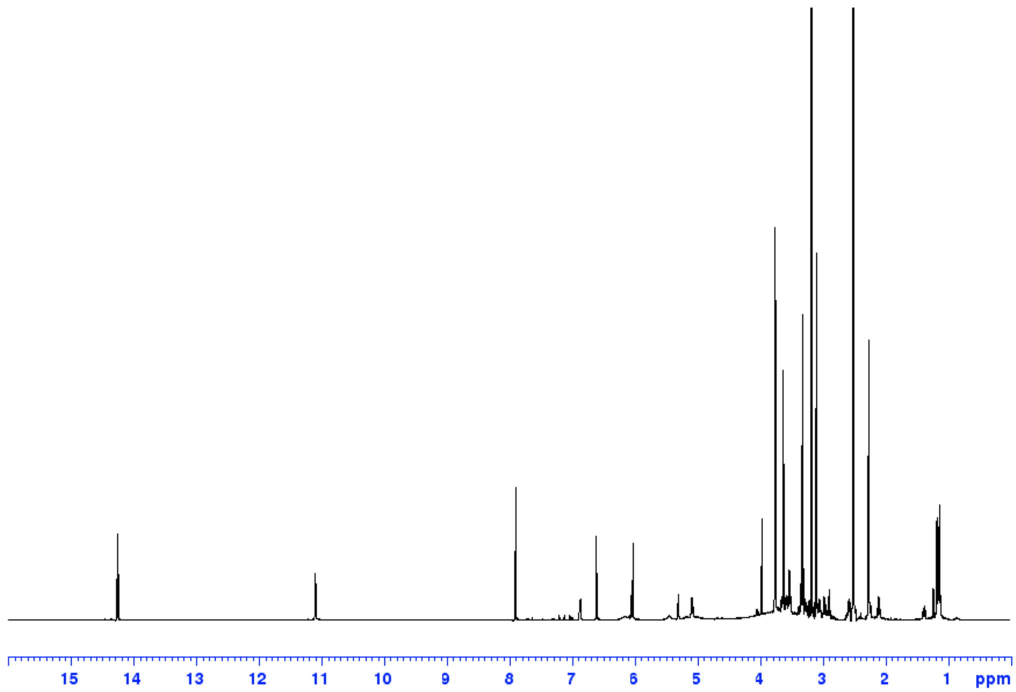
Supplementary figure 13. HSQC spectrum (DMSO-d_6, 500 MHz) of calixanthomycin A (**1**) at 100 °C.



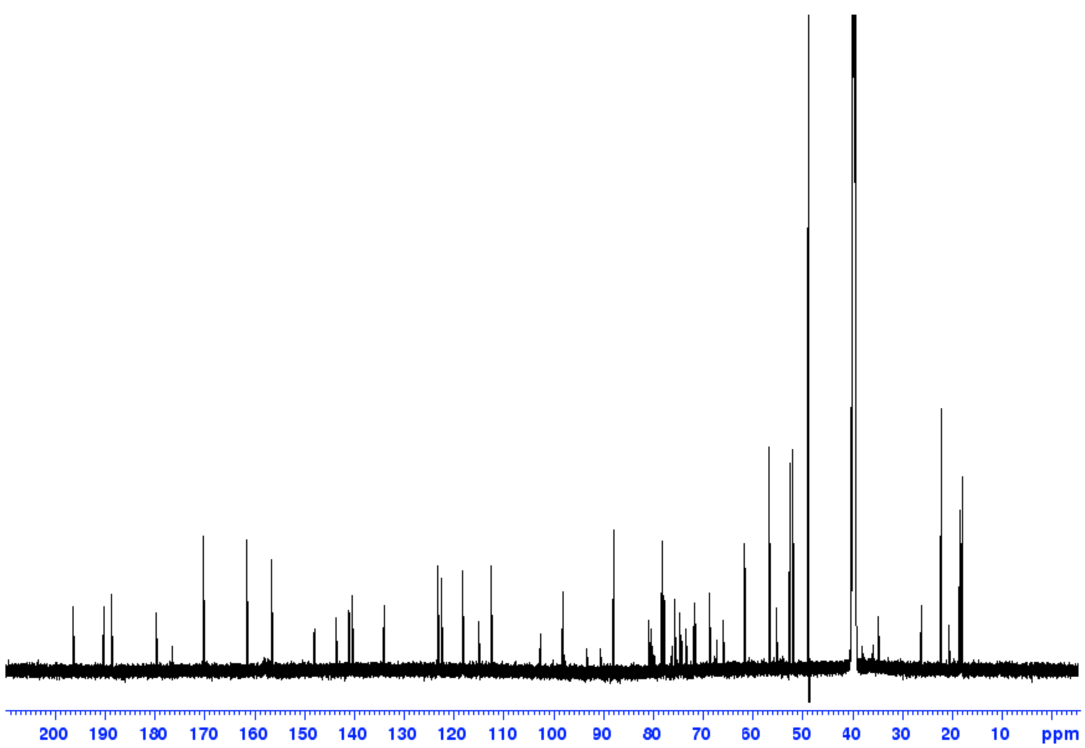
Supplementary figure 14. HMBC spectrum (DMSO-d_6, 500 MHz) of calixanthomycin A (**1**) at 100 °C.



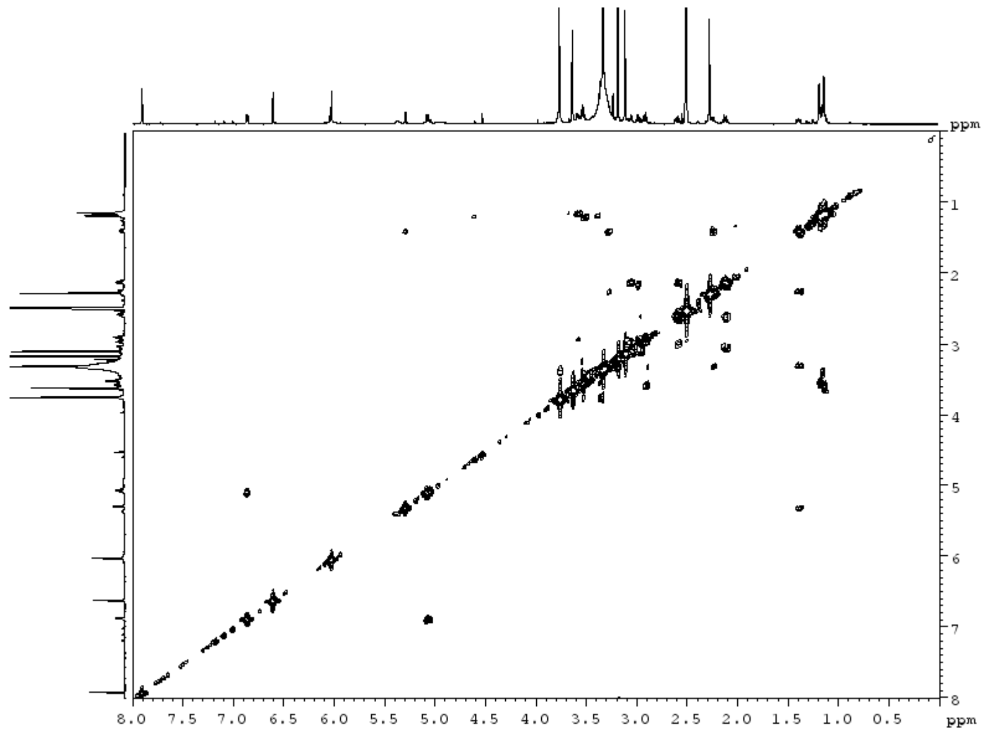
Supplementary figure 15. ROESY spectrum (DMSO-d_6, 500 MHz) of calixanthomycin A (**1**) at 100 °C.



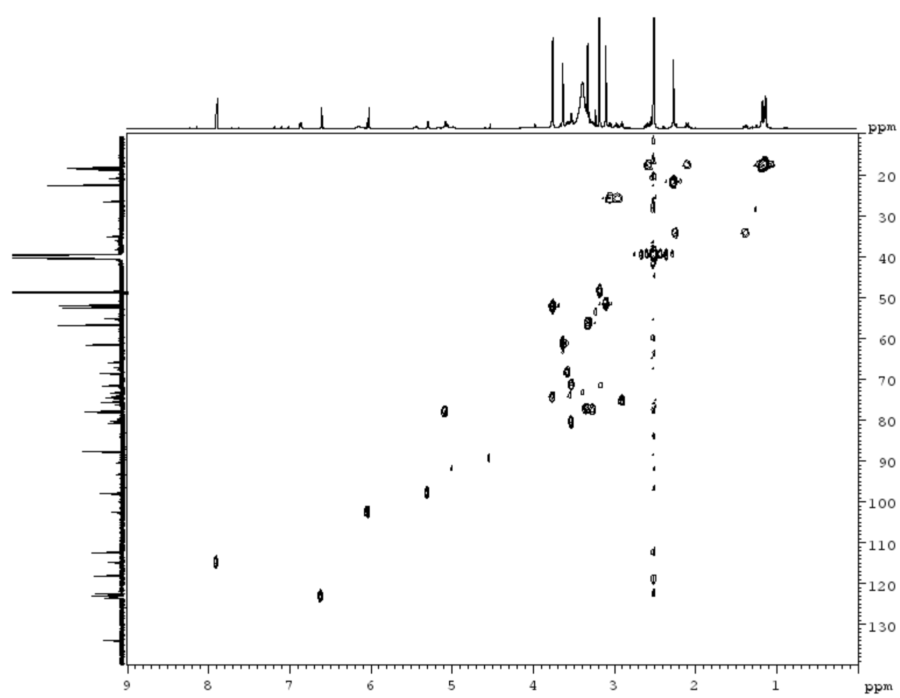
Supplementary figure 16. ^{1}H NMR spectrum (DMSO-d_6, 600 MHz) of arenimycin C (**2**).



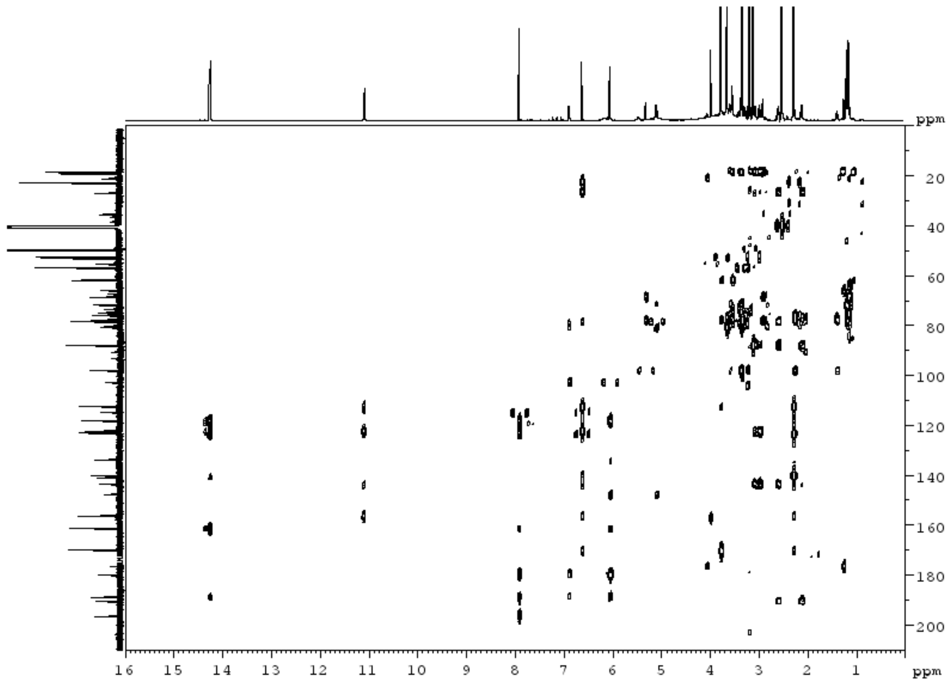
Supplementary figure 17. ^{13}C NMR spectrum (DMSO-d_6, 150 MHz) of arenimycin C (**2**).



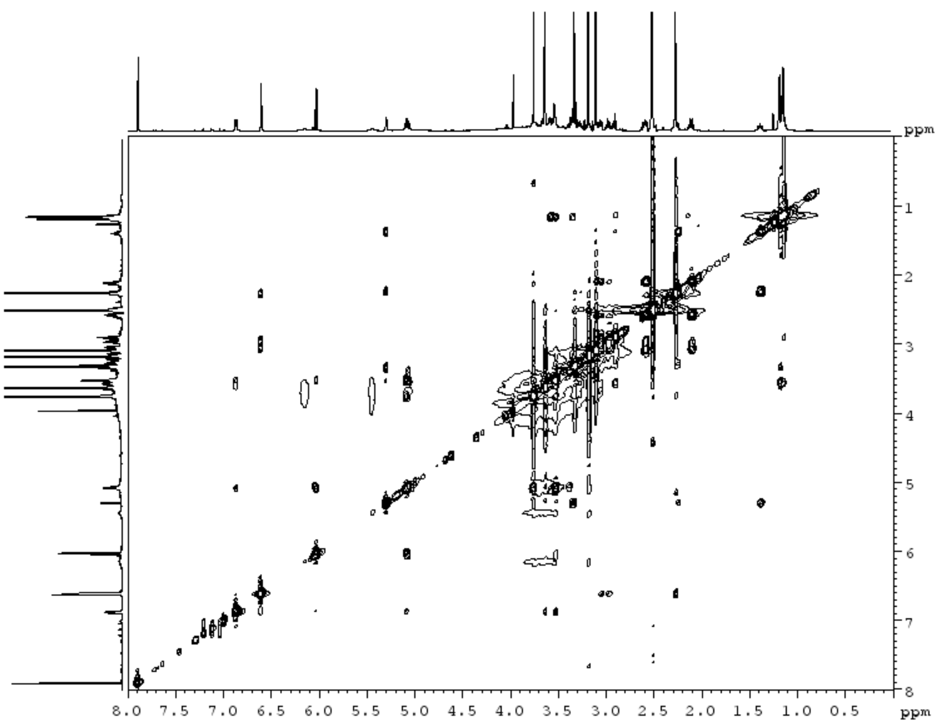
Supplementary figure 18. COSY spectrum (DMSO-d_6, 600 MHz) of arenimycin C (**2**).



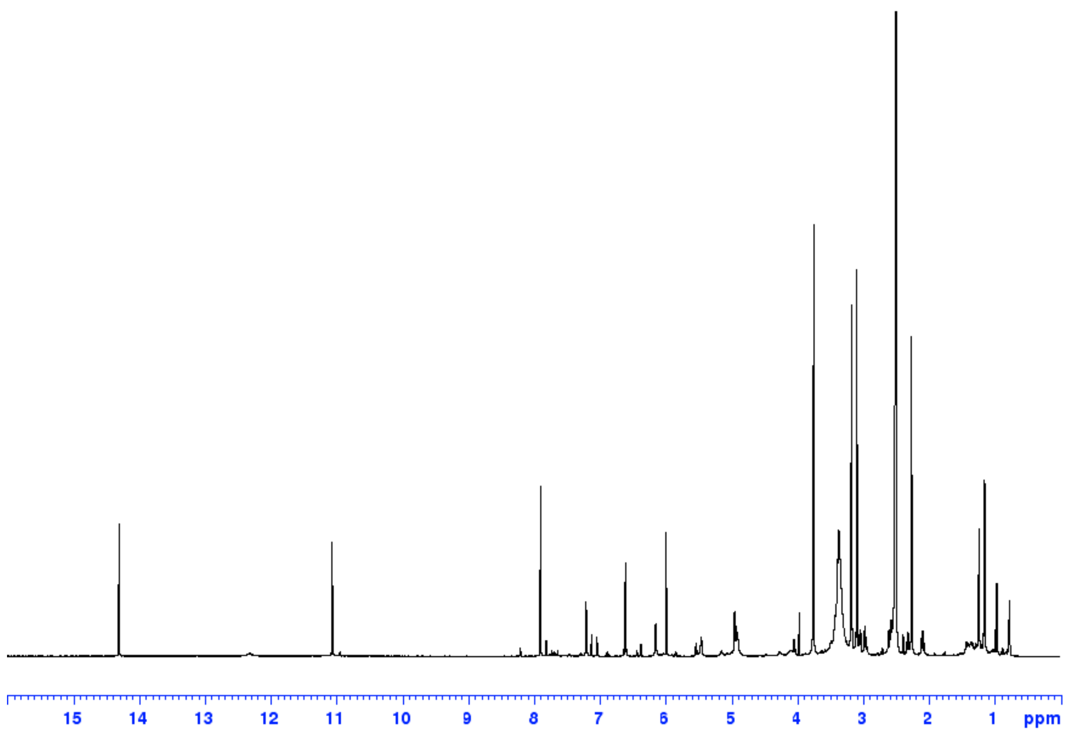
Supplementary figure 19. HMQC spectrum (DMSO-d_6, 600 MHz) of arenimycin C (**2**).



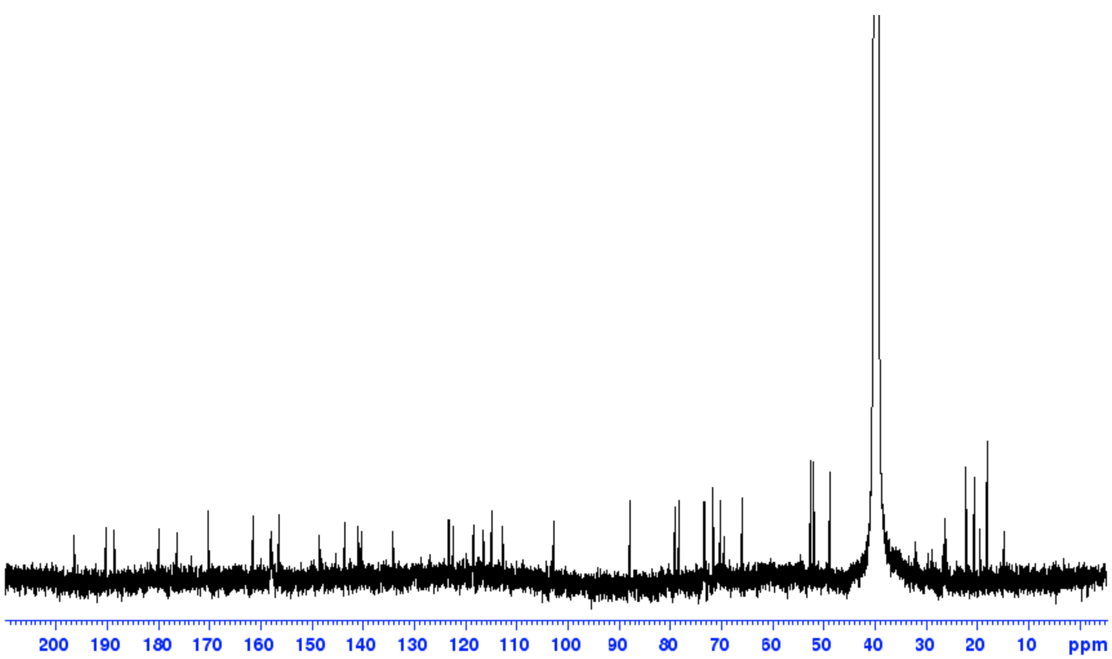
Supplementary figure 20. HMBC spectrum (DMSO-d_6, 600 MHz) of arenimycin C (**2**).



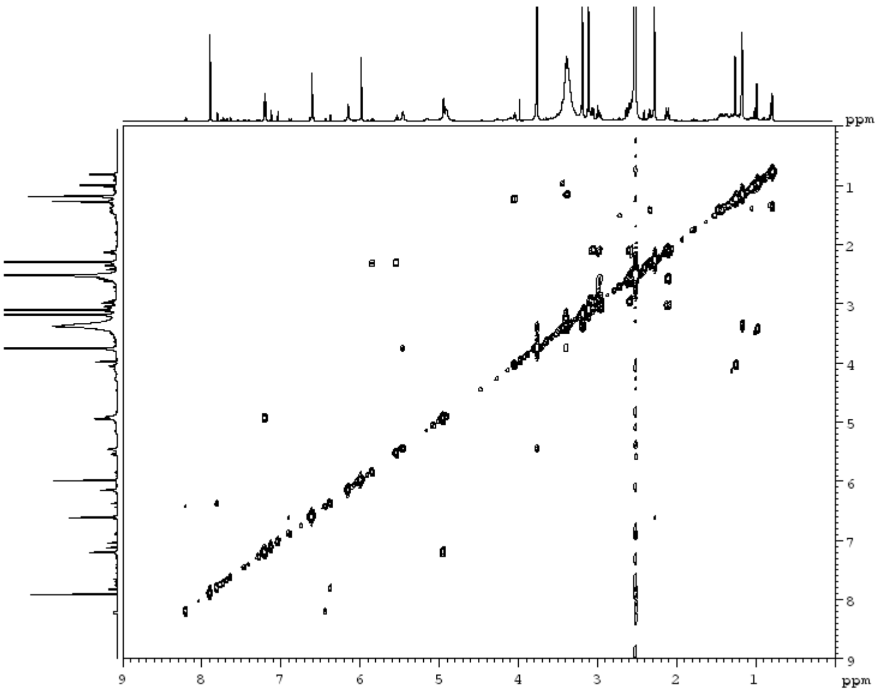
Supplementary figure 21. NOESY spectrum (DMSO-d_6, 600 MHz, mixing time: 600 ms) of arenimycin C (**2**).



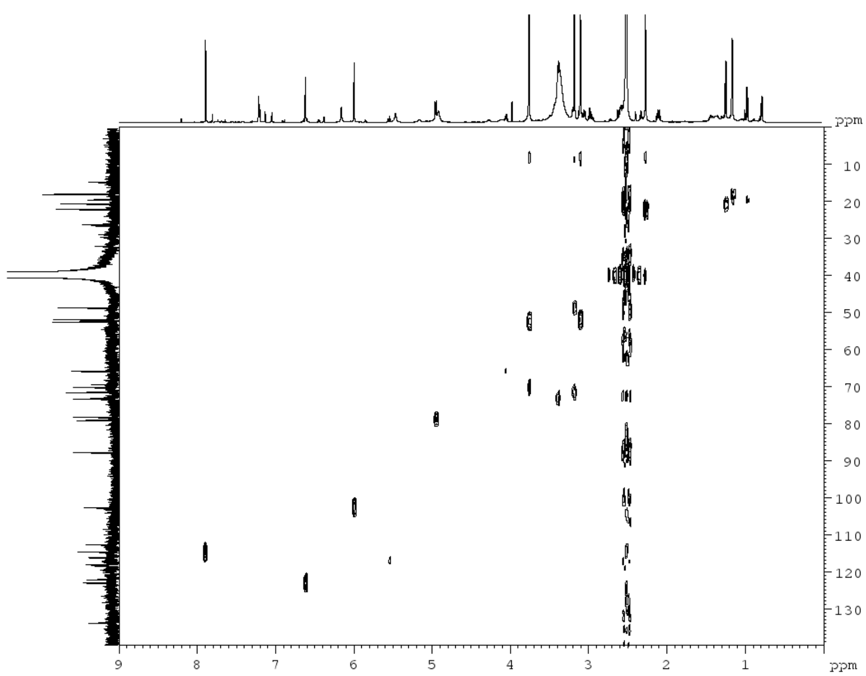
Supplementary figure 22. ^{1}H NMR spectrum (DMSO-d_6, 600 MHz) of arenimycin D (**3**).



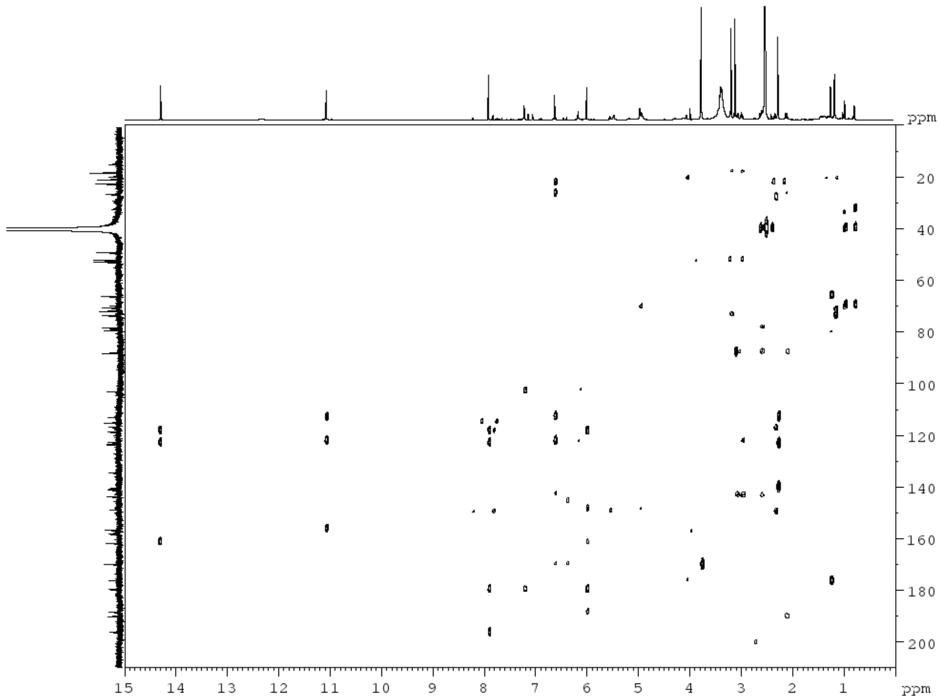
Supplementary figure 23. ^{13}C NMR spectrum (DMSO-d_6, 150 MHz) of arenimycin D (**3**).



Supplementary figure 24. COSY spectrum (DMSO-d_6, 600 MHz) of arenimycin D (**3**).



Supplementary figure 25. HMQC spectrum (DMSO-d_6, 600 MHz) of arenimycin D (**3**).



Supplementary figure 26. HMBC spectrum (DMSO-d_6, 600 MHz) of arenimycin D (**3**).